MISSING DATA METHODS: A REVIEW AND COMPARISON

By

Mary Ann Evans, PhD.

A Thesis

Presented to the Department of Public Health and Preventive Medicine

and the Oregon Health Sciences University

School of Medicine

in partial fulfillment of

the requirements for the degree of

Masters in Public Health

March 19, 2001

School of Medicine
Oregon Health Sciences University

CERTIFICATE OF APPROVAL

This is to certify that the MPH thesis of

Mary Ann Evans, PhD.

Has been approved

[redacted]

Professor in charge of thesis

[redacted]

Member

[redacted]

Member

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

This research project resulted from two years of postdoctoral training in health services research and studying for the Masters in Public Health. The faculty and research scientists at Oregon Health Sciences University and Kaiser Permanente Center for Health Research contributed in countless ways to my development and this project. I especially want to acknowledge the support from my committee members Drs. Donald K. Freeborn, John Mullooly, and Bentson McFarland. Dr. Freeborn deserves my added gratitude and thanks because he was also a member of my doctoral dissertation committee.

My husband Gary and our children Samantha, Annelise, and Zachary were loving, kind, and supportive during this journey. Without them I would not have had the courage or motivation to return to school and pursue yet another degree.

Above all I thank God for this accomplishment and the many blessings in my life and the world around me.

ABSTRACT

Missing data are a common and frustrating reality in research. The most serious threat of missing data is the introduction of systematic bias between observed and unobserved data that undermines validity. Theoretically and empirically supported methods for handling missing data exist, but are not commonly used because there is a lack of translations from technical publications. Many different methods exist for managing missing data, but it is unclear which method produces the most efficient and unbiased estimates.

The performance of listwise deletion, pairwise deletion, mean imputation, and Full Information Maximum Likelihood missing data methods were compared with different sample sizes, proportion of missing data, and different missing data mechanisms. Outcome variables for comparison included variance estimates, parameter estimates, error estimates, and variance accounted for in the dependent variable. The missing data methods produced similar results across conditions when data were missing completely at random. When data were missing at random and the sample was small, the Full Information Maximum Likelihood method produced consistent estimates while the listwise, pairwise, and mean methods produced mixed results. When data were not missing at random, the missing data methods produced similar results that varied significantly from the full sample.

CHAPTER 1

INTRODUCTION

Would the results be the same if no data were missing? Missing data are a common and frustrating reality in health services research. If there are any ways in which data can be missing, they will be (Cohen & Cohen, 1983). The most serious threat of missing data is the introduction of systematic bias between observed and unobserved data that undermines validity. Other problems include loss of statistical power, loss of efficiency, loss of information, and complications in data handling.

Health services research is under intense scrutiny from the public, funding agencies, and other researchers. How missing data are managed or mismanaged is an area of vulnerability. Theoretically and empirically supported methods for handling missing data exist, but are not commonly used. The preponderance of research and discussion about missing data exists primarily among statisticians in journals such as *Journal of Applied Statistics, Biostatistics,* and *Statistics in Medicine*. The lack of translations from theoretical and technical articles makes it difficult for applied researchers to use contemporary missing data methods.

Many different methods and variations in methods exist for managing missing data, making it unclear which method, if any, produces the most efficient and unbiased parameter estimates. The growing body of literature on missing data lacks synthesis of theory and research findings. I completed a missing data literature review and offer a systematic and evidence based review. The review is organized into sections about theory and assumptions, brief descriptions and research evidence regarding missing data methods. To illustrate strengths and weaknesses selected missing data methods were compared using real data.

LITERATURE REVIEW

It is clear from the literature that theory has not had much influence on practice in the treatment of missing data (Arbuckle, 1996; Worthke, 2000). Ignoring missing data is an option that is exercised often in research even though it leads to serious bias (Roth, 1994). Heitjan (1997) incorrectly characterized missing data as a nuisance. Missing data is a crucial element of research design and analysis. Treating missing data like a nuisance can lead to bias that undermines results. Appropriately managing missing data is an opportunity to reduce bias, improve power, increase efficiency, and increase confidence in research findings.

The key words missing, missing data, missing values, incomplete data, and nonresponse were used to search relevant DIALOG databases: MEDLINE, HealthSTAR, PsycINFO, Sociological Abstracts, Economics Literature Index, ERIC, ABI/Inform, Applied Social Science Index, and Humanities. The review focused on the last five years with key references from prior years.

Prior to 1985 less than 10 articles per year were published about missing data. Publications steadily increased averaging more than 25 articles per year in the last five years. The growth of interest in missing data is attributable in part to newly available computer software such as NORM, MX, AMOS, and the SPSS missing data module that makes complex missing data methods more accessible.

Roth (1994) reviewed 75 articles to catalog the use and reporting of missing data methods in research. More than 70% of the articles failed to mention missing data and only 12% stated the missing data method used. Listwise deletion was used most frequently (53%) followed by pairwise deletion. Many authors reported findings in ways that obscured missing data problems. The results highlight the lack of attention and deficiencies in reporting and publishing about missing data in research.

Since 1995, more than 150 articles were published that focused on missing data. About 20% of them were theoretical or discussion papers and 80% provided empirical data. Among the empirical publications, 75% compared missing data methods, 19% focused on a single missing data method without comparison, and 6% assessed the missing data mechanism or pattern of missing. About 36% of the empirical articles focused on managing missing data due to dropout and the remainder focused on cross-sectional data. *Statistics in Medicine* published special issues about missing data in 1997 and 1998 for a total of 42 articles. The 1997 issue focused on incomplete covariates, two-stage designs, and dropouts. The 1998 issue focused on missing data in quality of life research in cancer trials. A summary of contemporary theoretical and statistical issues and findings follows.

THEORY AND ASSUMPTIONS

Data are missing because potential respondents are not contacted, refuse to participate, withdraw from the study, and/or because participants fail to provide responses to one or more variables. Researchers use different terms to describe missing data that can be confusing. Nonresponse is a general term that includes failure to contact, refusals, withdrawal, attrition, and dropout. Variable nonresponse occurs when participants provide responses to some items, but not to others. Total nonresponse and variable nonresponse may occur simultaneously or individually in any given study.

Little and Rubin (1987) pioneered missing data research. To facilitate understanding, their definitions of missing data mechanisms along with Arbuckle's (1996) illustration of the principles are provided. Missing values of a dependent variable can be missing completely at random (MCAR), missing at random (MAR) or not missing at random (NMAR). Under MCAR, whether a variable's data are observed or missing is not thought to affect its distribution, that is $P(Y|y \text{ missing}) = P(Y|y \text{ observed})$. MAR is a less restrictive condition and assumes that missing and observed distributions of the dependent variable are identical when conditioned on a set of predictor variables, that is $P= (Y|y \text{ missing}, X) = P(Y|y \text{ observed}, X)$. Missing data are NMAR if the response probability depends on the missing data. When missing is associated with part of the variance not explained by other variables in the model, the residual variance, data are NMAR (Neal, 2000). Missing mechanism theory and research focus on missing in the dependent variable and not on independent variables.

Arbuckle (1996) provides an excellent illustration of the principles. Suppose a survey contains one item about income and one about education. Everyone answers the question about education, but not about income. If responses and nonresponses to income are independent of both income and education, then the data are missing completely at random (MCAR). In this case, respondents to the income question are a random sample of all respondents. If highly educated respondents are less likely to answer the question about income, but the probability of reporting income is unrelated to income given education, then the data are missing at random (MAR). In this case, respondents are a random sample of each educational group and the covariate education must be included in parameter estimation. If people with the same education level are less likely than others to report their income and the nonresponse is related to income, then data are not missing at random (NMAR) and respondents are not a random sample. The fundamental point about NMAR data is that the statistical model must incorporate the missing mechanism and not just observed data.

Groves and Cooper (1998) illustrated the effects of missing data with graphs of frequency distributions of responders and nonresponders. When responders and nonresponders possessed similar dependent variable distributions there was little bias regardless of the response rate. When the response rate was high and nonresponders reported higher values on the dependent variable than respondents, the sample mean underestimated the population mean, but bias was relatively small because of the high response rate. When the response rate was low and there was a large difference between respondents and nonrespondents, the area under the nonresponse curve and bias were large. The problem is that researchers rarely know the magnitude of the difference between responders and nonresponders for multiple variables and time periods.

Direct examination of the missing patterns is necessary because no omnibus statistical procedure exists to determine whether data are MCAR, MAR, or NMAR. It is critical to determine the nature and distribution of missing data before deciding on a method for managing it. Researchers need to identify why data are missing, the patterns of missing data, how much data are missing by variable and by person, and the characteristics of responders and nonresponders. Empirical solutions for missing data are often presented without first addressing why and how data are missing (Lubeck, Pasta, Flanders & Henning, 2000; Smeding & De Koning, 2000; Huberman & Langholz, 1999; Corfec, Chevret & Costagliola, 1999).

Data that are NMAR are difficult to analyze because an explicit model of missing is required and researchers frequently lack enough information to develop a model. Although not always possible, the best way to address the problem of not knowing the mechanism of missing is to collect additional data from or about nonresponders (Graham, Hofer & Pancinin, 1994). Graham, Hofer, Donaldson, MacKinnon and Schafer (1997) provide a compelling argument and evidence that NMAR is relatively rare and when present generally has little effect on statistical conclusions. They argue that most causes of missing are measurable and can be incorporated into prediction models that reduce or eliminate bias.

Missing data mechanisms (MCAR, MAR, NMAR) influence parameter estimation. Modeling missing is not necessary when data are MCAR because the sample is a random sample of the population. When data are MAR, covariates that are related to missing in the dependent variable must be included to produce unbiased estimates. The missing data mechanism must be included in the model when data are NMAR.

Little and Rubin (1989-1990) describe two theoretical frameworks for managing missing data. The aim of the first framework is to "fix up" the data by replacing missing data and restoring the rectangular form of the data matrix. The data is then treated and analyzed as if it were complete. The second theoretical orientation is direct analysis of the incomplete data without any attempt to restore the rectangular shape of the data. Brief definitions and evaluation of missing data methods are provided.

MISSING DATA METHODS

Listwise deletion. Listwise deletion (LD) is a special case of "fixing up" the data where cases that contain any missing value are discarded. Sample moments (means, variances, and covariances) and parameter estimates are calculated from the complete cases. Listwise deletion produces unbiased parameter estimates if the data are MCAR, but seriously biased estimates if data are MAR or NMAR. The greatest drawback of LD is loss of statistical power because many cases are discarded. Even if per variable rate of missing is low few participants may have complete data for all variables (Schafer & Olsen, 1998). LD estimates are biased and inefficient (Arbuckle, 1996; Brown, 1994; Little & Rubin, 1987). The advantage of LD is complete data that statistical software easily accommodates.

Pairwise deletion. In pairwise deletion (PD) sample moments are calculated by excluding cases with missing values on one or both variables. The method uses all available data in the sense that every observed value enters into the calculation of the sample moments. Although sample moments represent different cases and unequal numbers of cases, parameter estimates are calculated as if the sample moments came from complete data. If data are missing, the likelihood function does not simplify and PD can't solve the unsimplified form. PD pretends that the data are complete (Arbuckle, 2001). With the pretense, pairwise deletion produces unbiased parameter estimates if the data are MCAR, but seriously biased estimates if data are MAR or NMAR (Graham & Hofer, 2000; Arbuckle, 1996). Although PD may produce unbiased covariance estimates, there is no guarantee that the estimates are matrix unbiased (Marsh, 1998). The covariance matrix may not be positive-definite and may not maximize any likelihood.

The choice of the sample size influences parameter estimates and chi-square tests (Marsh, 1998). PD does not provide a valid means to obtain parameter estimates, standard errors, or a method for testing hypotheses (Arbuckle, 1996). The primary advantage of pairwise deletion is the inclusion of all available data in analyses.

Imputation. Imputation is a family of missing data methods that shares the principle of fixing up the data by replacing it and restoring the rectangular form of the data matrix. Imputation methods differ in how they replace data and the number of imputations. Regression (stochastic or nonstochastic), mean replacement, weighting, Cold Deck Imputation, Hot Deck Imputation, nearby neighbor, Carry Last Observation Forward, substitution, and sensitivity analysis are imputation approaches to replacing missing data. Multiple imputation is a special case of imputation and is discussed in the next section on maximum likelihood.

Regression imputation replaces missing values with predicted values from a regression of variables with missing values on observed items. Stochastic regression adds a residual to regression imputation to reflect uncertainty in the predicted value. Regression imputation can inflate variance and covariance estimates.

Mean replacement is a special case of regression where missing values are substituted with the mean of the observed values. Weighting is related to mean imputation. The main objective of weighting is to reduce bias due to nonresponse by making each respondent represent a different fraction of the target population. If the design weights are constant in subclasses of the sample then weighting and mean imputation will produce the same population mean estimates, but not the same sampling variances. Mean substitution underestimates variances and covariances and suppresses standard error estimates. Researchers argue adamantly that mean substitution should never be used (Graham, Hofer & Piccinin, 1994).

Cold Deck Imputation replaces missing values with a constant value from an external source such as a previous study. Hot Deck Imputation replaces missing data by substituting values from similar participants in the study. Nearby neighbor imputation replaces missing data by substituting values from cases next to the case with missing data. In longitudinal research, the Carry Last Observation Forward method replaces missing data in follow-up with previous responses. Substitution replaces nonrespondents with alternate participants not selected originally. Substitutes may differ systematically from nonrespondents and should be treated as imputed values (Little & Rubin, 1987). Sensitivity analysis replaces variable nonresponses with extreme positive scores, extreme negative scores, and varying proportions of scores.

While imputation is intuitively appealing, a single imputed value cannot represent all of the uncertainty in the data and can distort the association between variables (Little & Schenker, 1995). Standard analyses applied to imputed datasets can overestimate the precision of parameter estimates, can suppress or inflate variance estimates, can lead to standard errors that are too small, p values that are artificially low, and higher than expected type I errors (Schafer & Olsen, 1998). There is no valid method for estimating error with imputation. To appropriately estimate errors with imputation methods bootstrapping or other simulation methods are necessary.

Maximum Likelihood (ML). It is important to distinguish between maximum likelihood as a parameter estimator and maximum likelihood models for missing data. Maximum likelihood can be used with any missing data method as an estimator of parameters. Using ML as an estimator does not resolve the problems associated with the missing data method (Arbuckle, 1996). For example, using ML as the estimator with listwise deletion does not restore the size of the original sample. The term maximum likelihood (ML) refers to the missing data model and not the estimation method in this paper. Excellent reviews of ML principles with incomplete data are provided by Arbuckle (1996), Neal (2000), and Worthke (2000).

Probability and likelihood are closely linked but distinct principles that are helpful in explaining maximum likelihood missing data methods. Neal (2000) provides an excellent and nontechnical discussion of the principles using a coin toss experiment with an outcome probability of $p = .5$ as an example. Likelihood tests whether the coin is biased by comparing the height of the curve at various levels of the parameter p. In the coin example the hypothesized probability $p = .5$ is compared to different experimental outcomes. The likelihood at the maximum for the experiment is then compared to the hypothesized likelihood that is equal to the probability.

Maximum likelihood is a family of missing data methods that share the principle of direct analysis of incomplete data without any attempt to restore the rectangular shape of the data. The assumption is that there is a known form of the likelihood. In a sense there is no difference between ML estimation for complete and incomplete data (Little & Rubin, 1987). The ML method has a strong heuristic appeal because one chooses the parameter estimate that makes the observed data seem most likely. For example, take a sample drawn independently from a distribution with a probability function, f (y; theta), and an unknown vector parameter theta. The ML estimate is that value of theta which maximizes the likelihood over the set of all possible values for theta.

Optimal properties of ML include consistency, asymptotic normality, and asymptotic efficiency. ML is consistent because the estimate converges in probability to the true value of the parameter. The ML estimate converges to a multivariate normal distribution with a covariance matrix that is the inverse of the information matrix (asymptotic normality). ML is asymptotically efficient because it produces the best asymptotically normal estimate in terms of variance in large samples (Eliason, 1993). Full Information Maximum Likelihood (FIML), Multiple Group Structural Equation Modeling (MGSEM), and Expectation Maximization (EM) are different model based ML missing data methods.

FIML estimates the likelihood function at the individual case level. A remarkable mathematical property of the individual likelihood method is that it offers a simple but powerful treatment of missing data. Individual likelihood does not use any imputation, but simply calculates the likelihood of the observations that are present. The overall likelihood is made up of the product of individual likelihoods that are based on different numbers of observed variables (Neal, 2000). Twice the negative log-likelihood of the data is calculated for each observation (Arbuckle, 1996; Neal, 2000). Unlike the pairwise delete method, FIML can solve the unsimplified form of the likelihood (Arbuckle, 2001). FIML provides unbiased estimates of standard errors without simulation.

With MGSEM the sample is divided into groups with identical missing data patterns. The multiple group model is tested by imposing equality constraints across parameters and using FIML to estimate model parameters. The fit of the model is evaluated with and without equality constraints. MGSEM produces unbiased estimates of standard errors without simulation (Allison, 1987).

The Expectation Maximization (EM) algorithm is based on the idea that one makes guesses about the missing data values and uses those values to estimate sums, sums of squares, and cross products in the expectation (E) step. These sufficient statistics are then used to calculate the covariance matrix in the maximization (M) step. The updated covariance matrix is then used to estimate missing values in the next E step. The process continues until the elements of the covariance matrix stop changing (Graham & Hofer, 2000). Parameter estimates are virtually identical to FIML estimates. EM does not estimate standard errors because the likelihood function is solved at the moment level and not the individual case level.

Multiple imputation is an extension of EM. With multiple imputation multiple data sets are created through simulation by replacing missing data with plausible values. Parameter estimates are calculated and compared for each simulated data set. When there are many different patterns of missing, it is difficult to devise an effective imputation plan for each variable that takes into account data available for all other variables and the associations between the variables. Variables used in imputation may themselves be subject to missing data (Brick & Kalton, 1996). Significant effort and analytic resources are needed to simulate several data sets, simulate errors, repeat analyses, and average results.

FIML, MGSEM, and EM produce equivalent parameter estimates. Under MCAR parameter estimates are unbiased. ML estimates are less biased and more efficient under MAR and NMAR than ad hoc methods (Arbuckle, 1996; Worthke, 2000; Neal, 2000). FIML, MGSEM, and EM have different advantages and disadvantages. The primary advantage of EM with multiple imputation is that the rectangular form of the data is restored and can be analyzed and shared as if complete. FIML and MGSEM do not restore the rectangular form of the data. EM requires significant programming resources compared to FIML. EM requires bootstrapping or simulation for error estimates while FIML and MGSEM produce error estimates simultaneously with parameter estimates. EM is limited by the number of missing data patterns while FIML is not. MGSEM is limited when there is no solution for very small subgroups.

Bayesian. In a Bayesian framework missing data are considered random variables. Missing data are assigned a prior distribution expressing probabilistic uncertainty or degrees of belief about their values. Model parameters are also regarded as random variables. Inference about parameters is based on the marginal posterior distribution resulting from an integration with respect to missing data (Richardson & Leblond, 1997). Data augmentation (DA) is a Bayesian approach that shares some features of EM. DA is similar to EM because it iterates between an imputation step where missing data are simulated given a covariance matrix and a posterior step where the covariance matrix parameters are simulated given values of the data. DA converges differently than EM. When DA converges the distribution of the covariance matrix stops changing compared to EM where the elements of the covariance matrix stop changing. Misspecification of the prior distribution biases parameter estimates.

Another group of missing data methods evolved for managing dropout in longitudinal studies. The methods share features of ad hoc and ML missing data methods, but comparative studies and evidence are sparse.

Longitudinal missing data methods. Little (1993, 1994, 1995) described a general class of models termed pattern mixture models for missing longitudinal data. In these models participants are divided into groups based on missing data patterns for outcome variables. Pattern mixture models characterize the distribution of an individual's data conditional on time of dropout and then average over dropout times to determine treatment effects (Little, 1994). Mixed models handle unbalanced data by deleting fixed-effects parameters for missing levels, but do not delete missing level combinations for random effect parameters. Pattern mixture models can be used to model data that is NMAR. The pattern mixture approach can be used with random effects models, structural equation models, or Generalized Estimating Equation (GEE) based models.

GEE is a quasi-likelihood approach for modeling nonresponse. The method estimates parameters in a mean model with correlated data without making distributional assumptions. Weighted Estimating Equations (WEE) adapts the GEE approach and uses only complete case data with weights. Joint Estimating Equations (JEE) is an extension of WEE with joint probability added.

Random effects regression models are also called variance component models, hierarchical linear models, multilevel models, random coefficient models, two-stage models, mixed models, and unbalanced repeated measures models (Hedeker & Gibbons, 1997). In random effect models, responses may not be measured at the same time points or for the same number of time points. Participants who are missing at a given time point are not excluded from analysis. The model assumes that available data are representative of that participant's deviation from the average trend lines observed for the whole sample. The model estimates the participant's trend across time with available data augmented by the time trend for the whole sample and effects of all covariates in the model. Maximum likelihood is used for parameter estimation. It is assumed that for a

Ad hoc procedures are never better than recommended ML procedures and can yield results that are extremely biased (Graham & Hofer, 2000; Arbuckle, 1996).

Models that address some features of data missingness, even if they do not describe precisely the underlying mechanisms, may provide acceptable approaches for inference and estimation (Espeland, Craven, Miller & D'Agostino, 1999). As discussed, the preponderance of research and discussion about missing data exists primarily in statistical journals. The lack of translations from theoretical and technical articles and lack of synthesis of theory and research findings make it difficult for applied researchers to use contemporary missing data methods. Selected missing data methods are compared using real data to illustrate strengths and weaknesses.

CHAPTER 2

METHODS: COMPARISON STUDY

A 1992 survey of health care workers served as the data source for the comparison study (Freeborn, Pope & Schmoldt, 1993). The response rate for the survey was 75% with 4,999 respondents. For illustrative purposes the comparison was limited to 22 survey items in a multivariate regression model.

As a first step, I examined item nonresponse per respondent and found that 79 people did not provide responses to any demographics and 19 people did not provide responses to 10 or more items. These 2% did not provide enough information to estimate responses, were effective nonresponders, and were deleted to create a sample with 4,901 respondents.

In a second step, I examined missing data patterns and found a range of 1% to 7.8% missing per variable. Missing data frequencies are listed in Table 1. I created dummy coded missing indicators, analyzed sample correlation matrices and found nonsignificant associations with missing indicators.

Table 1 Frequency of Missing Data

	No. Miss	% Miss
Tenure	56	1%
Education	383	7.8%
Age	198	4%
Commitment	132	2.6%
Job control	63	1%
Job satisfaction	216	4.4%
Supervision	332	6.7%
Health Status	58	1%

Next, a multiple regression model was constructed and tested where age, education, tenure, organizational commitment, job control, job satisfaction, and quality of supervision predicted health status. The dependent variable, health status, was a single item variable with a five-item Likert type scale with responses ranging from very poor to excellent. The frequency distribution is displayed in Table 2. The dependent variable was positively skewed.

Job satisfaction and supervision were each eight item measures with Likert type response scales. Scale scores were calculated by summing across items and dividing by the number of items. Scale scores were coded missing if any item was missing. The remaining constructs were each measured with a single item. Detailed descriptions of the measures and psychometric properties are given in Freeborn et al. (1993).

Table 2
Health Status Frequency Distribution

Excellent	1764	36%
Good	2680	55%
Fair	368	8%
Poor	31	1%

I compared listwise deletion, pairwise deletion, mean imputation, and Full Information Maximum Likelihood (FIML) missing data methods. Listwise deletion, pairwise deletion, and mean imputation were selected because they are the most commonly used missing data methods and serve as a point of reference. FIML was selected in order to compare a maximum likelihood method with common methods. FIML produces the same parameter estimates as EM, but unlike EM calculates valid error estimates without simulation. Another reason for selecting FIML is the need for published research. Of the more than 150 published articles about missing data, less than 10 used or discussed FIML. Because data were cross sectional, no longitudinal method was selected. The data were analyzed with the Statistical Package for the Social Sciences (SPSS) and the Analysis of Moment Structures (AMOS) module of SPSS. AMOS was selected because the FIML method for missing data is an available process.

The performance of the missing data methods was compared under conditions of varying sample size, proportion of missing data, and different missing data mechanisms.

Performance criteria included differences in sample size, means, variances, parameter estimates, standard errors, and variance accounted for in the dependent variable. The "true" estimates were unknown because real data were used. Because the response rate was high, the sample was very large, and there was relatively little missing data, the full sample was treated as "true" for comparison of the different missing data methods.

A random number generator was used to select samples of 1000 and 500 respondents from the full sample. Results for samples of 1000 are not reported because there was little or no difference in comparison to the full sample. Findings are available from the author. A random number generator was used to create datasets with 4%, 10%, 20%, and either 30% or 40% missing in the dependent variable. For condition of data MCAR, missing in the dependent variable was randomly distributed across respondents. For the MAR condition, missing was constrained to occur among participants with education greater than or equal to five. For the NMAR condition, missing was constrained to occur among participants with education greater than or equal to five and health status equal to one.

CHAPTER 3

RESULTS

The full sample was used as the comparison group because the data were not simulated and the "true" estimates were unknown and because there was little difference in estimates across missing data methods and conditions. The results are organized into sections comparing the performance of the different missing data methods in terms of sample size, variance estimates for the dependent variable, parameter estimates, and variance accounted for in the dependent variable in comparison to the full sample. Within each section the missing data methods are compared for the different sample sizes, proportion of missing data, and missing data mechanism.

SAMPLE SIZE DIFFERENCES

Sample sizes varied widely among the missing data methods. Mean imputation and FIML used the full sample for each condition. PD produced unequal samples and LD significantly reduced the sample size. The default sample size in SPSS was used for the PD method. The reduction of the sample size reduced power and influenced sample moment and parameter estimates. Variations in means, variances, parameter estimates, errors, and prediction of the dependent variable were due in part to differences in sample sizes. See Table 3 for sample sizes.

MEAN SCORE DIFFERENCES

Mean scores on health status varied little across methods, proportion missing, sample size variations, or missing data mechanism. The mean score on health was 1.7 and varied between 1.6 and 1.9 when 20% or more of the data were missing.

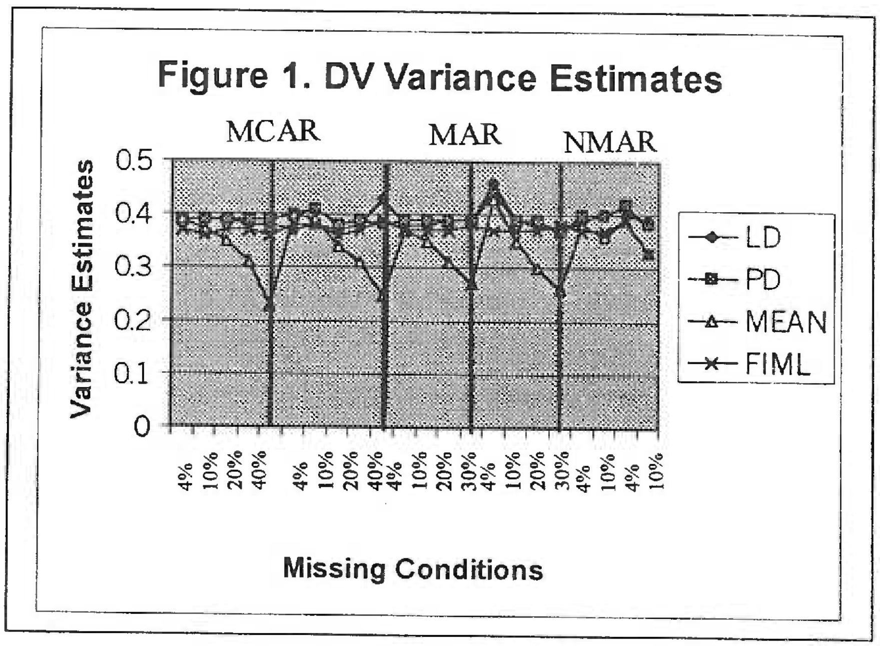
Table 3. Sample Size Differences

	Proportion Missing in the Dependent Variable				
	Full	4%	10%	20%	40%
Full MCAR:					
LD	3812	3703	3456	3065	2300
PD	4517-4877	4517-4877	4401-4877	3922-4877	2941-4877
MEAN/FIML	4901	4901	4901	4901	
500 MCAR:					
LD	390	377	349	320	231
PD	458-495	458-494	445-494	399-494	295-494
MEAN/FIML	500	500	500	500	500
Full MAR:					
LD		3688	3433	3020	2554
PD		4517-4845	4408-4845	3927-4845	3377-4845
MEAN/FIML		4901	4901	4901	4901
500 MAR:					
LD		374	352	310	265
PD		451-497	450-497	393-494	352-494
MEAN/FIML		500	500	500	500
Full NMAR:					
LD		3690	3444		
PD		4517-4845	4412-4845		
MEAN/FIML		4901	4901		
500 NMAR:					
LD		371	343		
PD		449-494	406-494		
MEAN/FIML		500	500		

VARIANCE ESTIMATE DIFFERENCES

Dependent variable variance estimates for each method, missing data mechanism, proportion missing, and sample size are displayed in Figure 1. As expected, mean imputation suppressed variance estimates across the different sample sizes, amount of missing data, and missing mechanisms. Variance estimates were very similar for LD, PD and FIML methods. LD and PD methods departed somewhat from FIML estimates when the sample size was 500 and the proportion of missing data exceeded 20%.

The vertical lines in the Figure 1 indicate the missing mechanism condition and sample size. The first section is the full sample with data MCAR. The second section is the 500 sample with data MCAR. The third section is the full sample with data MAR. The fourth section is the 500 sample with data MAR. The last section is data NMAR with the full sample and sample of 500 respectively. Within the vertical lines are the results for the different proportions of missing data. The layout for Figure 1 is the same for all subsequent figures.



DIFFERENCES IN PARAMETER ESTIMATES

In the full sample, significant independent variables included tenure, education, commitment, satisfaction, and supervision. Parameter estimates and standard errors for the significant independent variables comparing the different missing data methods, sample sizes, proportion missing, and missing conditions are listed in Table 4.

Table 4. Parameter and Error Estimates

	MCAR					MCAR				
	N=4901	4%	10%	20%	40%	N=500	4%	10%	20%	40%
Tenure:										
LD	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.02 (.01)*	0 (.03)	0.01(.03)	0.01(.03)	0.01(.03)	0.008 (.04)
PD	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.02 (.01)*	0.02 (.01)*	0.01 (.03)	0.002(.03)	0.02(.03)	0.02(.03)	0.003 (.04)
MEAN	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.02 (.01)*	0.01 (.01)*	0.01 (.03)	0.001(.03)	0.02(.02)	0.01(.03)	0.004 (.02)
FIML	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.02 (.01)*	0.02 (.01)*	0.01 (.03)	0.001(.03)	0.02(.03)	0.02(.03)	0.005 (.03)
Education:										
LD	0.11 (.01)*	0.11 (.01)*	0.12 (.01)*	0.11 (.01)*	0.11 (.01)*	0.07(.03)*	0.07 (.03)*	0.06(.03)	0.08 (.04)*	0.07 (.04)
PD	0.11 (.01)*	0.12 (.01)*	0.12 (.01)*	0.11 (.01)*	0.11 (.01)*	0.09(.03)*	0.09 (.03)*	0.08(.03)*	0.09 (.04)*	0.08 (.05)
MEAN	0.11 (.01)*	0.11 (.01)*	0.11 (.01)*	0.09 (.01)*	0.06 (.01)*	0.08(.03)*	0.08 (.03)*	0.07(.03)*	0.07 (.03)*	0.05 (.03)
FIML	0.11 (.01)*	0.12 (.01)*	0.12 (.01)*	0.11 (.01)*	0.12 (.01)*	0.09(.03)*	0.08 (.03)*	0.08(.03)*	0.08 (.04)*	0.08 (.04)*
Commit:										
LD	0.01 (.01)	0.01 (.01)	0.01 (.01)	0.01 (.01)	0.02 (.01)	0.01 (.03)	0.01 (.03)	0.02(.04)	0(.04)	.05 (.04)
PD	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.04 (.01)*	0.04 (.03)	0.04 (.03)	0.03(.03)	0.04(.03)	0.04 (.05)
MEAN	0.03 (.01)*	0.03 (.01)*	0.02 (.01)*	0.02 (.01)*	0.02 (.01)*	0.04 (.03)	0.04 (.03)	0.02(.03)	0.03(.03)	0.03 (.02)
FIML	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.04 (.01)*	0.04 (.03)	0.04 (.03)	0.03(.03)	0.03(.03)	0.05 (.04)
Satisfact:										
LD	0.1 (.02)*	0.1 (.02)*	0.1 (.02)*	0.1 (.02)*	0.09 (.03)*	0.13 (.06)*	0.14 (.06)*	0.11(.06)	0.11 (.07)	0.06 (.06)
PD	0.1 (.02)*	0.1 (.02)*	0.1 (.02)*	0.1 (.02)*	0.09 (.02)*	0.11 (.06)	0.11 (.06)	0.11(.06)	0.11(.06)	0.06 (.06)
MEAN	0.1 (.02)*	0.1 (.02)*	0.09 (.02)*	0.09 (.02)*	0.05 (.01)*	0.1 (.06)	0.1 (.06)	0.09(.05)	0.08 (.05)	0.03 (.04)
FIML	0.1 (.02)*	0.1 (.02)*	0.1 (.02)*	0.1 (.02)*	0.09 (.02)*	0.12 (.06)*	0.13 (.06)*	0.13(.06)*	0.12(.06)	0.05 (.07)
Super:										
LD	0.04 (.01)*	0.04 (.01)*	0.03 (.01)*	0.03 (.01)*	0.04 (.02)*	0.01 (.04)	0.09 (.04)	0.02 (.04)	0.02 (.04)	0.006 (.05)
PD	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.02 (.01)*	0.03 (.01)*	0.02 (.04)	0.01 (.04)	0.03 (.04)	0.05 (.04)	0.03 (.05)
MEAN	0.03 (.01)*	0.03 (.01)*	0.02 (.01)*	0.02 (.01)*	0.02 (.01)*	0.01 (.03)	0.003 (.03)	0.02 (.03)	0.04 (.03)	0.007 (.03)
FIML	0.04 (.01)*	0.04 (.01)*	0.04 (.01)*	0.02 (.01)*	0.04 (.01)*	0.02 (.04)	0.02 (.04)	0.03 (.03)	0.05 (.04)	0 (.04)

* denotes significant parameter estimates

Table 4 (continued). Parameter and Error Estimates

	MAR 4%	10%	20%	30%	MAR 500	10%	20%	30%	NMAR4%	10%	NMAR 500	10%
Tenure:												
LD	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.05 (.03)	0.07 (.03)*	0.05 (.03)	0.01 (.04)	0.03 (.01)*	0.03 (.01)*	0.008 (.03)	0.006 (.03)
PD	0.02 (.01)*	0.03 (.01)*	0.03 (.02)*	0.03 (.01)*	0.03 (.03)	0.07 (.03)*	0.06 (.03)*	0.03 (.03)	0.03 (.01)*	0.03 (.01)*	0.01 (.03)	0.007 (.03)
MEAN	0.02 (.01)*	0.02 (.01)*	0.02 (.01)*	0.02 (.01)*	0.03 (.03)	0.06 (.02)*	0.05 (.02)*	0.02 (.02)	0.02 (.01)*	0.02 (.01)*	0.01 (.03)	0.008 (.02)
FIML	0.02 (.01)*	0.02 (.01)*	0.03 (.01)*	0.03 (.01)*	0.02 (.01)*	0.02 (.01)*	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.01 (.03)	0.004 (.02)
Education:												
LD	0.11 (.01)*	0.12 (.01)*	0.11 (.01)*	0.11 (.02)*	0.12 (.03)*	0.06 (.04)	0.14 (.04)*	0.07 (.05)	0.13 (.01)*	0.18 (.01)*	0.09 (.04)*	0.16 (.03)*
PD	0.11 (.01)*	0.12 (.01)*	0.12 (.01)*	0.08 (.01)*	0.12 (.04)*	0.04 (.03)	0.13 (.04)*	0.03 (.04)	0.13 (.01)*	0.18 (.01)*	0.08 (.04)*	0.15 (.03)*
MEAN	0.11 (.01)*	0.1 (.01)*	0.07 (.01)*	0.04 (.01)*	0.12 (.03)*	0.03 (.03)	0.08 (.03)*	0.02 (.03)	0.12 (.01)*	0.17 (.01)*	0.08 (.03)*	0.15 (.03)*
FIML	0.11 (.01)*	0.12 (.01)*	0.12 (.01)*	0.1 (.01)*	0.11 (.01)*	0.12 (.01)*	0.12 (.01)*	0.1 (.01)*	0.13 (.01)*	0.18 (.01)*	0.07 (.03)*	0.15 (.03)*
Commit:												
LD	0.01 (.01)	0.02 (.01)	0.03 (.01)*	0.01 (.01)	0.08 (.04)*	0.1 (.04)*	0.03 (.04)	0.04 (.04)	0.01 (.01)	0.07 (.01)	0.01 (.04)	0.05 (.04)
PD	0.03 (.01)*	0.03 (.01)*	0.01 (.01)	0.03 (.03)*	0.09 (.04)*	0.1 (.03)*	0.06 (.03)	0.03 (.04)	0.03 (.01)*	0.02 (.01)*	0.002 (.03)	0.02 (.03)
MEAN	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.02 (.01)*	0.09 (.03)*	0.1 (.03)*	0.05 (.03)	0.02 (.02)	0.03 (.01)*	0.02 (.01)*	0.005 (.03)	0.02 (.03)
FIML	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.03 (.01)*	0.02 (.01)*	0.02 (.01)*	0.007 (.03)	0.03 (.03)
Satisfact:												
LD	0.1 (.02)*	0.09 (.02)*	0.11 (.02)*	0.13 (.02)*	0.17 (.07)*	0.03 (.06)	0.01 (.07)	0.16 (.08)	0.1 (.02)*	0.1 (.02)*	0.12 (.07)	0.21 (.06)*
PD	0.09 (.02)*	0.09 (.02)*	0.11 (.02)*	0.12 (.02)*	0.13 (.07)	0.004 (.06)	0.05 (.06)	0.14 (.07)	0.1 (.02)*	0.1 (.02)*	0.1 (.06)	0.15 (.06)*
MEAN	0.09 (.02)*	0.08 (.02)*	0.08 (.02)*	0.08 (.02)*	0.12 (.06)	0.08 (.05)	0.03 (.05)	0.09 (.05)	0.09 (.02)*	0.08 (.02)*	0.09 (.06)	0.12 (.05)*
FIML	0.1 (.02)*	0.09 (.02)*	0.11 (.02)*	0.12 (.02)*	0.1 (.02)*	0.09 (.01)*	0.11 (.02)*	0.12 (.02)*	0.1 (.02)*	0.1 (.01)*	0.09 (.06)	0.17 (.06)*
Super:												
LD	0.04 (.01)*	0.04 (.01)*	0.03 (.01)*	0.01 (.02)	0.03 (.04)	0.11 (.04)*	0.03 (.04)	0.04 (.05)	0.04 (.01)*	0.03 (.01)*	0.06 (.04)	0.02 (.04)
PD	0.04 (.01)*	0.04 (.01)*	0.03 (.01)*	0.01 (.01)	0.04 (.04)	0.08 (.04)*	0.003 (.04)	0.01 (.04)	0.03 (.01)*	0.03 (.01)*	0.06 (.04)	0.01 (.04)
MEAN	0.04 (.01)*	0.03 (.01)*	0.02 (.01)*	0.01 (01)	0.04 (.04)	0.08 (.03)*	0.006 (.03)	0.006 (.03)	0.03 (.01)*	0.03 (.01)*	0.06 (.03)	0.01 (.03)
FIML	0.04 (.01)*	0.04 (.01)*	0.03 (.01)*	0.01 (.01)	0.04 (.01)*	0.04 (.01)*	0.03 (.01)*	0.01 (.01)	0.03 (.01)*	0.03 (.01)*	0.06 (.04)	0.002 (.04)

* denotes significant parameter estimates

In general, power decreased when the sample was smaller and the proportion of missing data increased. Standard errors tended to increase with smaller samples and resulted in fewer significant independent variables. For illustrative purposes, estimates for education and satisfaction across conditions are displayed in Figures 2-5 and are discussed in detail.

Education was a significant predictor of health status across methods and proportion of missing data when data were MCAR except when the sample was 500 and the proportion of missing was 40%. In this condition LD, PD, and Mean parameter estimates were not significant. Parameter estimates ranged from .11 to .05. When the proportion of missing exceeded 10%, the Mean method suppressed parameter estimates.

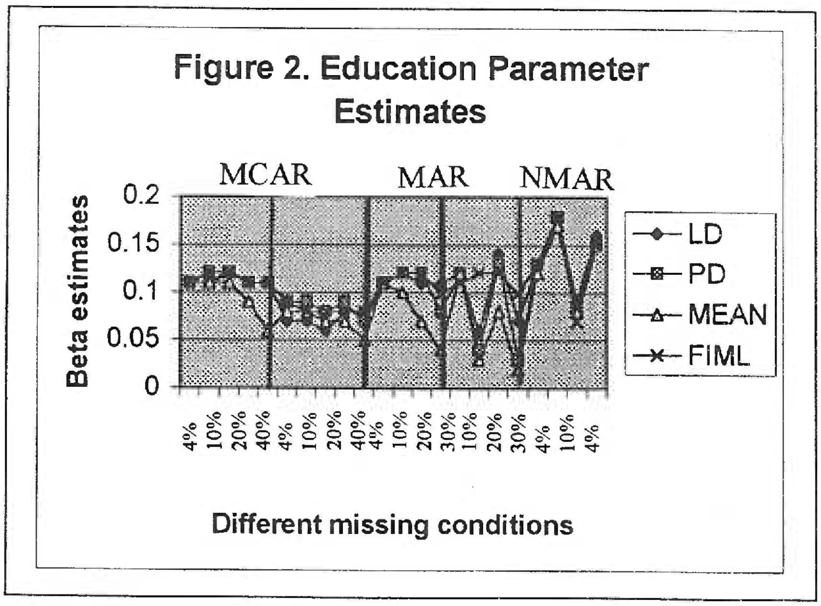
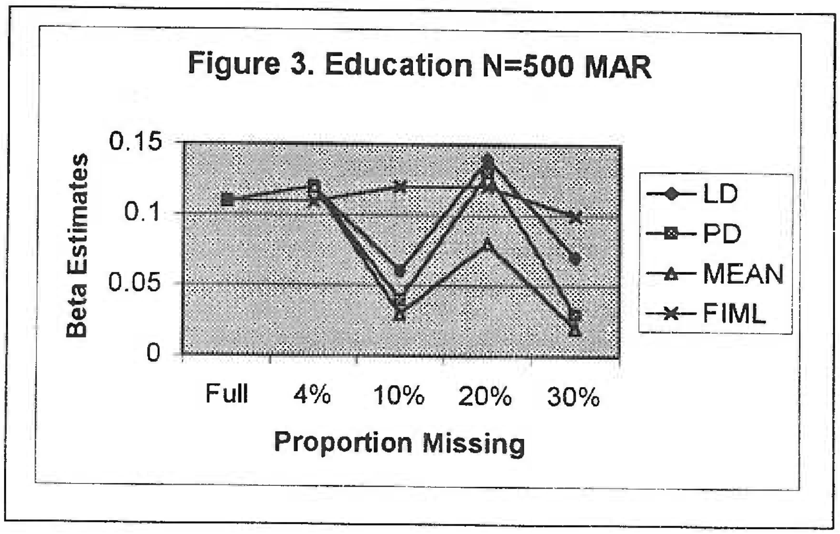


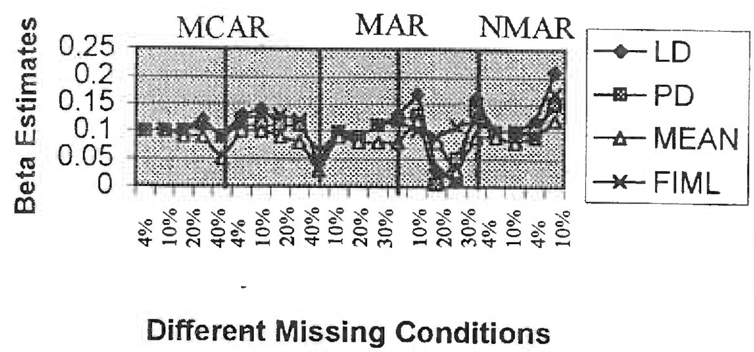
Figure 2. Education Parameter Estimates

Parameter estimates for education were significant across missing data methods when data were MAR and the full sample was used, but not when the sample was 500. The Mean method suppressed the estimate of education when the proportion of missing exceeded 10%. When data were MAR and the sample size was 500, the FIML method produced consistent and significant parameter estimates across proportions of missing data. LD, PD, and Mean methods estimates for education fluctuated and error estimates increased. See Figure 3 for parameter estimates when the sample was 500 and data were MAR. Parameter estimates and errors for education were similar across missing data methods when data were NMAR.



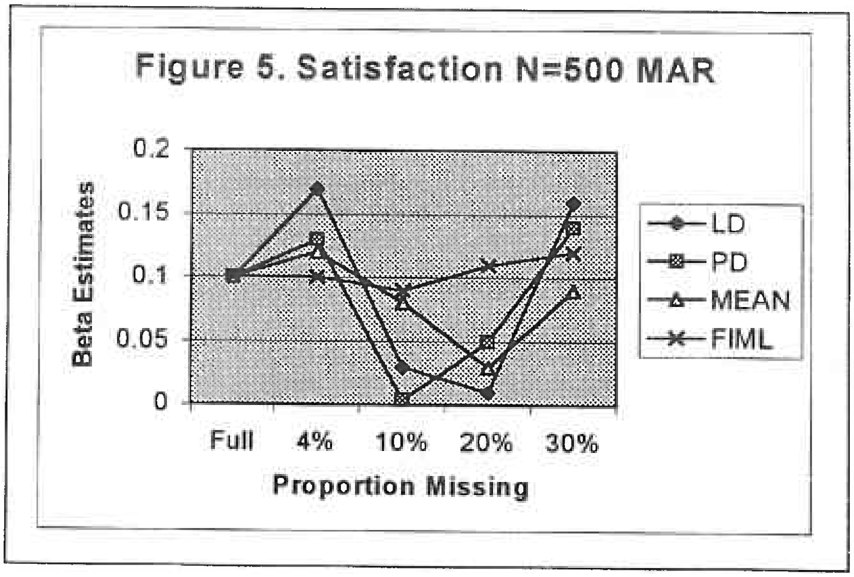
Satisfaction was a significant predictor of health status across methods and proportions of missing data when data were MCAR and the full sample was used. Parameter estimates ranged from .09 to .10. When the sample size was 500, parameter estimates were significant for FIML until the proportion of missing exceeded 10%. Parameter estimates were significant for LD until the proportion of missing exceeded 4%. PD and Mean methods did not result in significant parameter estimates for satisfaction when the sample size was 500. Parameter estimates ranged from .14 to .03. The standard errors of the parameter estimates increased dramatically from .01 to .02 when the full sample was used to .04 to .07 when the sample was 500.

Figure 4. Satisfaction Parameter Estimates



Parameter estimates for satisfaction were significant across methods and proportions of missing data when data were MAR and the full sample was used. When data were MAR and the sample size was 500, the FIML method produced consistent and significant estimates across proportions of missing data. LD, PD, and Mean methods estimates for satisfaction varied and error estimates more than doubled. Error estimates increased from .02 in the full sample to a range of .02 to .08 in the sample of 500. See Figure 5 for parameter estimates when the sample was 500 and data were MAR. Parameter estimates and errors for satisfaction were similar across missing data methods when data were NMAR.

None of the missing data methods performed well when data were NMAR. When data were NMAR and the full sample was used, parameter estimates for education were overestimated across missing data methods. When data were NMAR and the sample size was 500, parameter estimates for tenure, education, commitment, and satisfaction were underestimated and supervision was overestimated. When the proportion of missing data increased, parameter estimates for tenure and supervision were underestimated and



parameter estimates for education and satisfaction were overestimated in comparison to the full sample model.

Although parameter estimates and errors were sometimes similar across missing data methods, caution is warranted for interpreting LD, PD, and Mean estimates, because there is no valid method for producing the error estimates. Bootstrapping or resampling would be required.

VARIANCE ACCOUNTED FOR DIFFERENCES

Variance accounted for in the dependent variable for each method, proportion missing, sample size, and missing mechanism are displayed in Figure 6. Estimates were similar across LD, PD, and FIML methods under MCAR. The Mean method suppressed variance accounted for in the dependent variable.

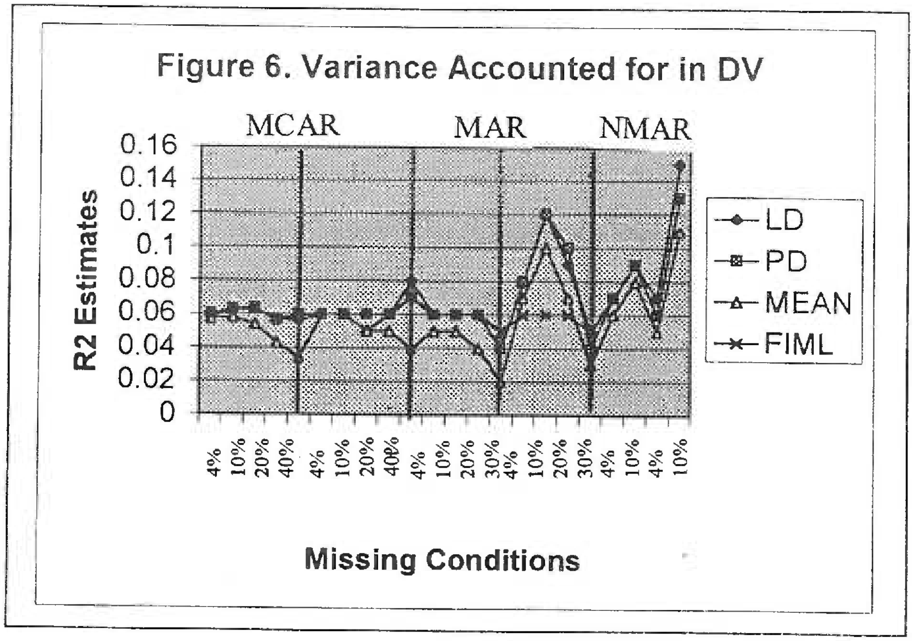


Figure 6. Variance Accounted for in DV

Estimates were similar for LD, PD, and FIML methods when data were MAR and the full sample was used and the Mean method suppressed estimates. When the sample was 500 LD, PD, and Mean methods overestimated variance accounted for. FIML provided consistent variance accounted for estimates. See Figure 7 for variance accounted for in the sample of 500 and data MAR. When data were NMAR, variance accounted estimates were similar across methods, but varied from the full sample model.

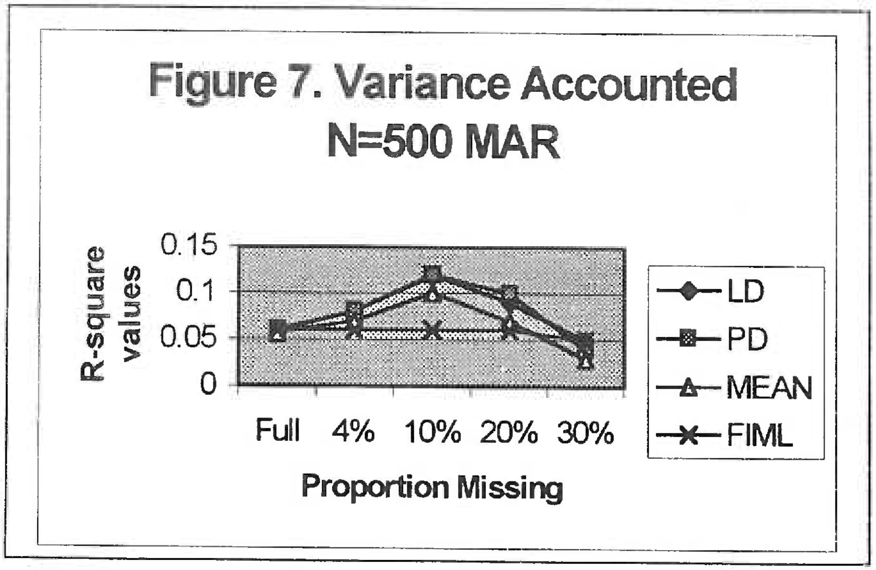


Figure 7. Variance Accounted
N=500 MAR
R-square values
0.15
0.1
0.05
0
Full
4%
10%
20%
30%
Proportion Missing
LD
PD
MEAN
FIML

CHAPTER 4

DISCUSSION

The strengths of the comparison study included a very large sample, a high response rate, and very little missing data. The large sample permitted comparison of the missing data methods while avoiding estimation problems associated with statistical power. The high response rate and very little missing data across variables and participants reduced potential nonresponse bias. There was little difference in estimates for the full sample and samples of 1,000 across methods and conditions. To better explore how the missing data methods performed under less ideal conditions, the sample size was reduced, the proportion of missing data was increased, and the missing data mechanism was varied.

A shortcoming of the comparisons was the lack of multiple simulated data sets. Multiple simulated datasets would reduce sampling error. The single sample design used in this study highlights the difficulties in research studies where a single sample of data is collected. A single random sample of the population protects against a researcher biasing the selection of participants, but not against differences in participant responses. While the study samples were selected completely at random, there were some differences between the samples of 500 and the full sample.

Sample size, means, variances, parameter estimates, standard error of parameter estimates, and variance accounted for in the dependent variable were compared for each missing data method and condition. The actual sample size varied widely for PD and LD missing data methods. FIML and Mean methods captured the full sample. Smaller samples reduced power to detect significant differences and contributed to variation in parameter estimates, errors, and prediction of the dependent variable. As expected, the mean score on health varied little across missing data methods whether the sample was small, the proportion of data missing was large, or the missing data mechanism was MCAR, MAR, or NMAR.

Variance estimates for health status were very similar for LD, PD, and FIML methods, while the Mean method suppressed variance. Variance accounted for in the dependent variable were similar across LD, PD, and FIML methods when data were MCAR. When the sample was 500 and data were MAR, the FIML method was the only method that produced consistent estimates. The results confirm findings from other studies that Mean imputation produces biased and inefficient estimates even when data are MCAR. Any advantage gained by retaining the full sample with mean imputation is lost in biased and inefficient parameter estimates.

When data were MCAR parameter estimates were similar for LD, PD, and FIML missing data methods. The Mean method tended to suppress parameter estimates. When data were MAR and the sample was 500, FIML produced significant and consistent parameter estimates in comparison to the full sample. LD, PD, and Mean methods overestimated or underestimated parameter estimates in comparison to the full sample. When data were NMAR, the missing data methods produced similar parameter estimates that varied from the full sample.

FIML appears to combine the advantages of PD and Mean missing data methods. From the PD method, FIML captures the unbiased pairwise covariance estimates, but avoids the pitfalls of PD by adopting a statistically valid method of estimating parameters and errors. From the mean method, FIML captures the entire sample, but avoids suppression of estimates. Another advantage of the FIML method that should not be underestimated is that the method produces parameter and error estimates in a single step without simulation, bootstrapping, or creating multiple data sets. FIML assumes the likelihood function is known.

The comparison study results provide preliminary support for an alternative method for testing the missing data mechanism. As discussed, there is no omnibus statistical test for determining whether missing data are MCAR, MAR, or NMAR. As an alternative, one could initially test for the missing data mechanism by comparing estimates produced by LD, PD, Mean, and FIML methods. Equivalent estimates would provide evidence that missing data were MCAR.

The best way to manage missing data is to avoid it. The next best course of action is to engage in efforts that minimize the effects of missing data and collect information about nonresponders. Without collecting additional data, the causes of missingness are unknown and uncertain. The better a researcher is able to account for missingness, the stronger the argument that important causes of missingness are measured and accounted for in prediction models.

The comparison study suffered limitations. For example, little was known about nonresponders. Even though the response rate was very high, it is possible that nonresponders differed significantly from responders. Freeborn et al. (1993) compared survey demographics such as occupation and gender with organizational demographics and found similar distributions for the study population. Although the comparison does not guarantee that there were no differences between responders and nonresponders, it provided some evidence that nonresponders and responders shared similar characteristics.

As discussed, another limitation was the use of single samples for comparing methods. More complex simulation studies are needed to address bias fully and compare the performance of the missing data methods. I selected single samples of 1,000 or 500 and randomly deleted data for each condition. Multiple samples with data deleted multiple times for each sample are needed to assess bias more accurately. The study also lacked comparison of other missing data methods such as regression, EM, multiple imputation, and Bayesian methods.

More research is needed to explore fully the performance of different missing data methods with real data under conditions of data MAR and NMAR. More than five years passed since Roth (1994) reviewed publications to assess the treatment of missing data. A more current review is needed to determine whether current missing data research improved the treatment and reporting of missing data in publications.

Returning to the question posed at the beginning of the paper, would the results be the same if no data were missing? The results would be the same if data were MCAR and Mean imputation was not used. Results would be nearly the same if the sample size was large, the response rate was high, and very little data was missing. The results would not be the same if data were MAR and LD, PD, or Mean missing data methods were used. The results would not be the same if data were NMAR whether a LD, PD, Mean or FIML missing data method was used.

REFERENCES

Allison, P. (1987). Estimation of linear models with incomplete data. In C.C. Clogg (ed.) *Sociological Methodology*. San Francisco: Jossey-Bass.

Arbuckle, J. (1996). Full information estimation in the presence of incomplete data. In G.A. Marcouldies and R.E. Schumacker (eds.) *Advanced Structural Equation Modeling*. Mahwah, NJ: Lawrence Erlbaum Publishers

Arbuckle, J. (2001). Personal communication.

Brick, J. & Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, 5, 215-238.

Brown, C. (1994). Efficacy of the indirect approach for estimating structural equation model with missing data: A comparison of five methods. *Structural Equation Modeling: A Multidisciplinary Journal*, 1, 287-316.

Cohen, J. & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Ass.

Corfec, E., Chevret, S. & Costagliola, D. (1999). Visit-driven endpoints in randomized HIV/AIDS clinical trials: Impact of missing data on treatment difference measured on summary statistics. *Statistics in Medicine*, 18, 1803-1817.

Eliason, S. (1993). *Maximum Likelihood Estimation Logic and Practice*. Newbury Patk, CA: SAGE.

Evans, M. (2000). Do missing data matter and what do you do about it in health services research. *Proceeding of the Association for Health Services Research.*

Espeland, M., Craven, T., Miller, M. & D'Agostino, R. (1999). 1996 Remington Lecture: Modeling multivariate longitudinal data that are incomplete. *Annals of Epidemiology*, 9, 3, 196-205.

Freeborn, D. K., Pope, C.R., & Schmoldt, R. (1993). Predictors of job satisfaction and psychological well-being among health care workers. Proceedings International Sociological Association Mental Health and Illness Research Committee. Rome, Italy.

Graham, J. & Hofer, S. (2000). Multiple imputation in multivariate research. In T. Little & K. Schnabel (eds) *Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Graham, J., Hofer, S., Donaldson, S., MacKinnon, D., and Schafer, J. (1997). Analysis with missing data in prevention research. In K. Bryant, M. Windle & S. West (eds.) *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research.* Washington D.C.: American Psychological Association.

Graham, J., Hofer, S. & Piccinin, A. (1994). Analysis with missing data in drug prevention research. In L. Collins & L Seitz (eds.) *Advances in Data Analysis for Prevention Intervention Research (NIDA Monograph no. 142).* Washington D.C.: National Institute on Drug Abuse.

Groves, R. & Cooper, M. (1998). *Nonresponse in Household Interview Surveys.* New York: John Wiley & Sons, Inc.

Hedeker, D. & Gibbons, R. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2, 1, 64-78.

Heitjan, D. (1997) Annotation: What can be done about missing data? Approaches to imputation. *American Journal of Public Health*, 87, 4, 548-50.

Huberman, M. & Langholz, B. (1999). Application of the missing-indicator method in matched case control studies with incomplete data. *American Journal of Epidemiology*, 150, 12, 1340-1345.

Little, R. (1994). A class of pattern-mixture models for multivariate incomplete data. *Biometrika*, 81, 471-83.

Little, R. (1995). Modeling the dropout mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90, 1112-121.

Little, R. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88, 125-34.

Little, R. & Rubin, D. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Little, R. & Rubin, D. (1989-1990). The analysis of social science data with missing values. *Sociological Methods and Research*, 18, 292-326.

Little, R & Schenker, N. (1995). Missing data. In G. Arminger, C.C. Clogg & M.E. Sobel (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences.* New York: Plenum.

Lubeck, D., Pasta, D. Flanders, S. & Henning, J. (1999). Approaches to missing data inference results from CaPSURE. *Pharmacoeconomics*, 15, 2, 197-204.

Marsh, H. (1998). Pairwise deletion for missing data in structural equation models: Nonpositive definite matrices, parameter estimates, goodness of fit, and adjusted sample sizes. *Structural Equation Modeling*, 5, 22-36.

Neal, M. (2000). Individual fit, heterogeneity, and missing data in multigroup structural equation modeling. In T. Little & K. Schnabel (eds.) *Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Park, T. (1993) A comparison of the generalized estimating equation approach with the maximum likelihood approach for repeated measurements. *Statistics in Medicine*, 12, 18, 1723-32.

Richardson, S. & Leblond, L. (1997). Some comments on Misspecification of priors in Bayesian modeling of measurement error problems. *Statistics in Medicine*, 16, 203-213.

Roth, P. (1994). Missing data: A conceptual view for applied psychologists. *Personnel Psychology*, 47, 537-560.

Schafer, J. & Olsen, M. (1998). Multiple imputation for multivariate missing-data problems: A data analysts perspective. *Multivariate Behavioral Research*, 33, 4, 545-571.

Smeding, H.& De Koning, I. (2000). Frontotemporal dementia and neuropsychology: The value of missing values. *J Neurol Neurosurg Psychiatry*, 68, 6, 726-30.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T. Little & K. Schnabel (eds) *Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Zhao, L., Lipsitz, S. & Lew, D. (1996). Regression analysis with missing covariate data using estimating equations. *Biometrics*, 52, 4, 1165-82.